Turning a hobby into a job: How duplicated genes find new functions

Gavin C. Conant and Kenneth H. Wolfe‡*

Abstract | Gene duplication provides raw material for functional innovation. Recent advances have shed light on two fundamental questions regarding gene duplication: which genes tend to undergo duplication? And how does natural selection subsequently act on them? Genomic data suggest that different gene classes tend to be retained after single-gene and whole-genome duplications. We also know that functional differences between duplicate genes can originate in several different ways, including mutations that directly impart new functions, subdivision of ancestral functions and selection for changes in gene dosage. Interestingly, in many cases the 'new' function of one copy is a secondary property that was always present, but that has been co-opted to a primary role after the duplication.

Subfunctionalization
A pair of duplicate genes are said to be subfunctionalized if each of the two copies of the gene performs only a subset of the functions of the ancestral single copy gene.

**Division of Animal Sciences, University of Missouri–Columbia, 163B Animal Sciences Center, 920 East Campus Drive, Columbia, Missouri 65211-5300, USA.*
‡Smurfit Institute of Genetics, University of Dublin, Trinity College, Dublin 2, Ireland.
Correspondence to K.H.W.
e-mail: khwolfe@tcd.ie
doi:10.1038/nrg2482

In *The Origin of Species*[1], Darwin gave himself two challenges. First, he sought to demonstrate the wide kinship of life on Earth, an idea he termed descent with modification. Second, he contended that the agency responsible for directing those modifications was natural selection[1]. Darwin was fairly successful at convincing his contemporaries on the first point, but his argument for natural selection as a creative force was less readily accepted[2,3]. Although the power of selection in removing disadvantageous variants was clear, many doubted it could build wholly new structures from only random changes.

To understand the origins of a novel structure we need to answer two questions[3]: what is the genetic source of the novel structure? And how has that new structure become adapted to its function? In morphological studies, the first question is addressed by looking for homologies between the novel anatomical feature and existing structures in other taxa[3]. The implication is that new structures, including new genes, are not constructed *de novo* but are co-opted from existing ones (the 'Panda's Thumb' principle[4]). Such co-option often involves exaptation, in which the current, selectively beneficial function of something differs from its original, possibly non-adaptive role[5]. A good example is the discovery of feathers on non-avian dinosaurs[6,7]; it seems that feathers originally evolved for some other purpose and were later exapted for flight.

Co-option of an anatomical structure can occur either when the new function can coexist with the original function (feathers can be used for both flight and warmth), or when the original function becomes unnecessary in a new environment (fins with digits seem to have been co-opted in the move to dry land[8]). When the structure in question is a gene, however, it is generally assumed that a selective advantage for a novel function will rarely coincide with a relaxation of selection on the ancestral function. Thus, it was appreciated early on that the genetic 'backup' provided by gene duplication would be crucial to functional innovation[9].

The canonical work on the subject is Ohno's *Evolution by Gene Duplication*[10], which made the case for the importance of gene duplication and considered the various types of duplications and their potential for yielding novel functions. However, not all gene duplications involve functional innovation: in the years since Ohno's monograph we have realized that many (or even most) gene duplicates do not confer novel functions and are probably preserved in the genome by passive mechanisms such as subfunctionalization[11]. However, most evolutionary innovations in gene functions do seem to be associated in some way with gene duplication. Our interest here is the elucidation of the molecular origins of such new functions. Our key message is that the exaptation of pre-existing secondary functions is an important feature in gene evolution, just as it is in morphological evolution.

In the following sections we consider the two linked questions of how duplication can create genetic novelty by co-opting features from existing genes, and how selection optimizes that novelty. We use examples from

bacteria, yeast, vertebrates and plants. But first we review the data on gene duplication. Our focus is on how the evolutionary opportunities that gene duplication provides differ depending on whether the duplicate gene pair in question was formed by whole-genome duplication (WGD) or by single-gene duplication. We then discuss some examples of duplicate genes in which functional innovation has occurred and has been studied in detail. We highlight the importance of considering all the properties of a protein (not merely those previously shaped by selection) in understanding functional innovation. In some of these examples, we emphasize that we still do not know precisely how the novel function appeared: indeed, for genome-scale studies of duplicated genes, we might not know if novelty exists at all. Thus, we must be cautious in using genomic patterns of duplicate gene retention to infer the mechanisms by which novel functions appear. In the second part of this Review, we address this difficulty by considering the actions of both natural selection and genetic drift in the fixation and subsequent optimization of duplicate genes. We explore theoretical models and examples of the various ways in which duplicate genes can be preserved by natural selection, including the well known neofunctionalization and subfunctionalization models. Finally, we consider in detail two cases in which the biochemical mechanisms behind duplication-derived adaptations have been revealed. These examples are instructive reminders that our knowledge of how selection acts on duplicate genes is not yet exhaustive.

Which genes undergo duplication?

General trends. Individual gene duplications have been studied by geneticists for many years (for a historical perspective see REF. 9). One benefit of the genome-sequencing era is that it provided a complete and unbiased view of the landscape of duplicates present in each genome[12]. For our purposes, one of the most important outcomes of genome-wide studies is the identification of gene features that might allow or prohibit the fixation of a duplicate copy of that gene in the population. For instance, functional biases in the types of genes that survive in duplicate are known from a variety of organisms, including yeast, humans, insects and bacteria; unexpectedly high numbers of duplicated genes belong to categories such as transcription factors, kinases, and particular enzymes and transporters[9]. Genome-wide studies have also shown that duplication is favoured in a number of situations: among yeast and worm genes with below average rates of evolution[13]; among yeast genes that encode proteins that reside in less densely clustered parts of the protein-interaction network, implying that they have fewer pleiotropic constraints[14]; and among yeast genes with smaller fitness defects when knocked out[15], a pattern that seems to hold only for small-scale duplications (SSDs), not for WGD[16] (see below).

These four findings are both intriguing and confusing. For example, it is odd that, although slowly evolving genes tend to be found in duplicate more frequently, essential genes are not more likely to be duplicated despite the indirect association between lower rates of evolution and essentiality[17]. There are also several competing explanations for the biases in function and in rate of evolution among duplicates. But collectively these results are also intriguing because they suggest two potential reasons for the variation in duplicability among functional classes of genes. One possibility is that certain types of genes have biochemical features that allow them to be adapted easily to novel functions, such as the substrate promiscuity in the alcohol dehydrogenases (see below). A second possibility is that other types of genes might be particularly unlikely to undergo functional innovation via duplication, because the duplication has an immediate detrimental effect (so-called duplication-resistant genes[18]). Note the distinction: with the second possibility a gene might have a function that is potentially adaptable to a useful new function, but because the initial duplication is selectively disadvantageous the duplication cannot become fixed in the population and hence the adaptation is precluded. The notion that duplication might interfere with highly constrained cellular systems and hence be selectively unfavourable is termed the dosage-balance hypothesis[19]. It is supported by the observation that genes in sparse regions of the protein-interaction network[14] and with weaker knockout fitness defects[15] tend to have high rates of duplication. This is because genes with these properties will tend to have fewer of the dosage conflicts that could make duplication disadvantageous.

Single-gene duplications versus whole-genome duplications. Although the above studies were carried out using the yeast *Saccharomyces cerevisiae*, they either did not consider duplicate genes created by the WGD that took place in the evolutionary history of this species[15] or did not distinguish between duplicates that arose through SSD and WGD[13,14]. Not making this distinction is a potential concern as WGD and SSD can produce different kinds of adaptations[20]. However, comparisons between gene duplicates produced by WGD and by SSD also support the dosage-balance hypothesis.

The functional categories of duplicate genes retained after WGD bear unexpected similarities across diverse taxa. Functional classes that are over-represented among genes retained from the polyploidization events that took place in the ancestors of yeast, *Arabidopsis thaliana* and *Paramecium* spp. include ribosomal proteins and protein kinases[21–23]. In the plant and ciliate WGD duplicate gene sets, an excess of duplicated transcription factors was also observed[21,23]. Although this over-representation was not initially seen in yeast[22], it has been detected (G.C.C and K.H.W., unpublished observations) in a reanalysis that considered only proteins showing significant binding to DNA in a chromatin immunoprecipitation experiment[24].

In the *A. thaliana* genome, Maere *et al.*[25] found a pattern of anticorrelation between genes fixed in duplicate after single-gene duplication events and those surviving from WGDs — that is, genes in a functional category either tended to be duplicated by SSD and not by WGD, or vice versa. Similar results have been observed in yeast[20], with the additional observation that duplicates

Genetic drift
Random fluctuations through time in the allele frequencies of a population, caused by a sampling effect in small populations. Drift can overcome the effects of natural selection if the selective differences between alleles are small.

Neofunctionalization
A pair of duplicate genes in a population are said to be neofunctionalized if one of the two genes possesses a new, selectively beneficial function that was absent in the population before the duplication.

retained after WGD tend to diverge in expression more quickly, for a given level of coding-sequence divergence[26]. Duplicates produced by WGD also seem to share more protein interactions after duplication than do genes duplicated by SSD[26,27]. Finally, products of WGD are often highly expressed[22] and are more likely to show an overexpression phenotype or haploinsufficiency than other duplicates[20]. Strangely, although SSDs tend to be created from genes with smaller than average knockout fitness defects, enzymes that are retained in duplicate after WGD seem to have fitness defects at least as large as those for the genes that are not retained[16].

The above observations support the dosage-balance hypothesis because dosage balance is not, to a first approximation, altered by WGD. Thus, it is not surprising that polyploidization can lead to the retention of duplicates of genes whose dosage balance is potentially important (such as ribosomal proteins), whereas this class of gene is rarely duplicated by SSD. This idea implies that WGD events might allow certain evolutionary novelties to appear and be selected for that would have been unlikely to arise otherwise. Paradoxically, WGD might also allow innovation through changes in relative gene dosage. We have previously suggested that, because duplicate gene loss after WGD is common[28], it is possible for a WGD to allow all the genes in a pathway such as glycolysis to increase their relative dosages simultaneously, simply by remaining in duplicate as the rest of the genome shrinks[29]. This phenomenon of preservation of certain duplicate genes after WGD for reasons of dosage balance also might have larger implications in evolution. Freeling and Thomas[30] argue that the preferential preservation of members of gene complexes (broadly defined) after duplication represents an evolutionary 'drive' towards increased complexity in multicellular organisms.

Other than the question of dosage balance, one general principle that has emerged from genome-scale studies is that duplicate genes diverge most commonly in their regulation and least commonly in their biochemical function[20]. Indeed, two recent studies on the yeast WGD have suggested that duplicate genes retain substantial degrees of functional overlap for long periods (as indicated by shared protein-complex membership or by the observation of larger-than-expected double knockout fitness defects for a duplicate pair given the pair's single-gene knockout effects)[31,32]. These results reinforce the view that the majority of the gene duplications fixed in the yeast genome did not confer novel functions or have not subsequently diverged substantially in function. But it should be noted that the concepts of novelty and of the function of a gene are open to multiple definitions, and that other results — such as the finding that many WGD loci that are fixed in duplicate in one yeast species are single-copy in others[33] — point to a high frequency of neofunctionalization.

The nature of co-option. The question of which types of functions and genes are amenable to evolution following gene duplication has been addressed in depth by studying the co-option of gene functions at smaller experimental scales. Co-option often occurs using some evolved feature of the gene in question, although perhaps in a surprising way. For enzymes, this is most commonly the enzymatic activity. Alcohol dehydrogenase (ADH) enzymes have been co-opted for new functions in *Drosophila* (the *jingwei* gene, discussed below) and in yeasts. *S. cerevisiae* has a number of adaptations to a lifestyle that is based on glucose fermentation[29,34]. One such adaptation is a fixed ADH gene duplication in which the product of one of the duplicated genes is optimized for the conversion of acetaldehyde to ethanol, whereas the other product is optimized for the reverse reaction[35]. By reconstructing the ancestral enzyme, Thomson *et al.*[35] showed that it was optimized to convert acetaldehyde to ethanol. This finding implies that ethanol production was originally only used to regenerate enzyme cofactors under oxygen limitation, and that the ADH gene duplication helped produce the new phenotype of first excreting and then consuming ethanol. Duplication and later adaptive evolution has also allowed the co-option of an RNase gene for deriving nutrients from bacteria in a leaf-eating monkey[36].

Interestingly, co-option can also result in proteins being used for functions that are seemingly unrelated to their evolved role. Perhaps the classic example is the lens proteins of the eye, which have been co-opted from metabolic enzymes[37,38]. Another example is the antifreeze genes of Antarctic fishes, which were co-opted from repetitive sequences and parts of a duplicated trypsinogen gene[39]. Notably, initiation of a co-option event does not always require a gene duplication: co-option can occur first via gene sharing (one gene performing two functions[37]), with a subsequent duplication allowing the resolution of any adaptive conflicts produced by that gene sharing[40]. Another example of co-option concerns a large family of 'milk' protein genes in a viviparous cockroach species[41]. These proteins have sequence similarity to a secreted aphrodisiac protein from another cockroach species[42], and it has been suggested that both proteins might have originally been co-opted from an as yet unidentified ancestral lipid-binding protein[41].

Natural selection and co-option

We now move to the second question raised in this Review: how has natural selection acted to hone new gene functions introduced by duplication? We approach this question by considering the theoretical and experimental evidence for various mechanisms by which duplicate genes can be fixed (that is, reach a 100% frequency in a population) and preserved (that is, protected from loss by a selective advantage of genomes with two copies over mutants in which one copy has been deleted). This question is important given that, in the absence of other factors, gene duplications are expected to be either selectively neutral[43] or detrimental[44–46]. Thus, to explain the large numbers of long-lived gene duplications evident in genome sequences, we need to identify the compensating selective factors that have prevented their loss (which can occur either through genetic drift or through purifying selection)[47–49].

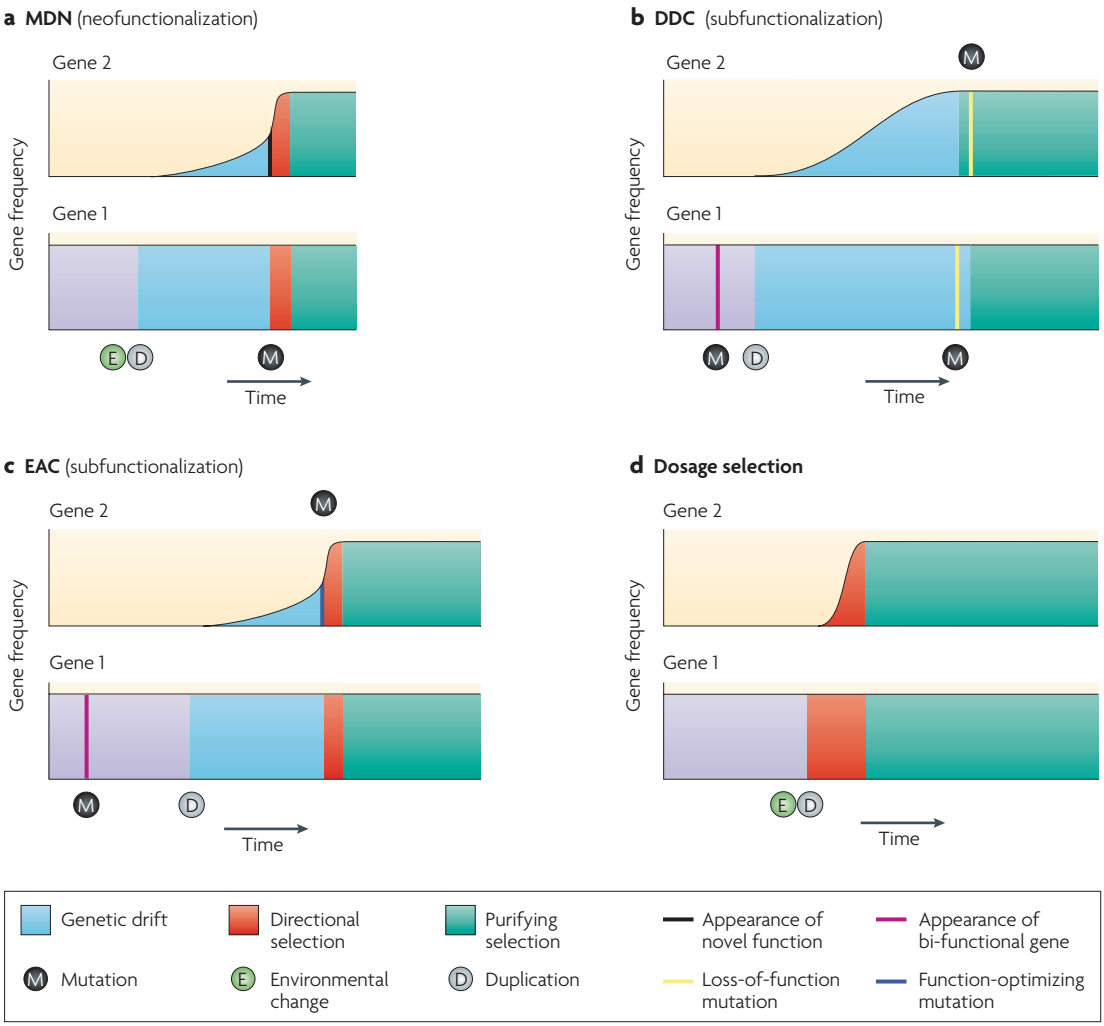


Figure 1 | **Potential fates of duplicate genes.** Each panel shows the outcome of a gene duplication event in terms of the population frequency of both the original gene and the resulting gene pair, which are present at separate genomic loci. Before the duplication, the locus containing the parental gene is fixed in the population. Models of how gene duplicates that arise in the population could first become fixed and then become preserved by purifying selection are shown. These examples illustrate the importance of the relative timing of events in the diverging fates of the duplicates. **a** | Mutation during non-functionality (MDN) neofunctionalization. In this case, environmental changes make the new gene function beneficial at point E. The duplication appears next, and it might or might not become fixed before the appearance of the neofunctionalizing mutation. **b** | Subfunctionalization by neutral degenerative mutations — the duplication, degeneration, complementation (DDC) model. Here both the fixation of the duplication and the subsequent subdivision of the ancestral functions occur through drift. Note that, for simplicity, we have combined the appearance and fixation of the degenerative mutations (yellow lines). After this process of subfunctionalization, neofunctionalization can also occur for one or both of the resulting gene copies. **c** | Subfunctionalization through escape from adaptive conflict (EAC). Unlike **b**, fixation of the duplication occurs by directional selection after a mutation in one copy (blue line) that optimizes that gene for one subfunction while making it less able to perform the other function. Again, after this process of subfunctionalization, neofunctionalization can also occur for one or both of the resulting gene copies. **d** | Duplication fixation through dosage selection. Here, environmental changes make increased dosage beneficial. Subsequent neofunctionalization is again a possibility for both duplicate gene copies.

To do this, we need to consider the three life stages of a duplicate gene pair: creation, fixation–preservation and subsequent optimization (FIG. 1). A duplication begins as a mutation in a single individual. The duplication will then either be lost from the population, or become fixed in the population either by natural selection in favour of it (FIG. 1a,c,d) or by genetic drift (FIG. 1b). Loss of a duplication is by far the most common fate, but one that will not be addressed in this Review because it does not give rise to new functions. In the case of fixation by genetic drift, the duplication also needs to become preserved to prevent future neutral loss. The final stage is the post-fixation modification of function, in which mutations conferring novel functions might appear and become fixed in the population by directional selection. Thus, we note that selection operates on duplicate genes in a number

Box 1 | Sequence evolution and neofunctionalization

With the availability of ever larger quantities of genome sequence data, there has been considerable interest in sequence-based methods of detecting functional changes[103–105]. Because neofunctionalization imparts a selective advantage to the possessors of the new trait, it is expected that the associated mutations will become fixed in a population more quickly than mutations that are fixed through genetic drift[106]. If a sufficient number of such favourable substitutions are fixed, they can elevate the non-synonymous substitution rate per non-synonymous site (K_A, also called d_N) to a value significantly larger than the synonymous rate (K_S, also called d_S). This result is not possible under genetic drift[107], meaning that testing the significance of $K_A > K_S$ is a test for directional selection[108]. Unfortunately, the converse is not true, because factors such as divergence time and variable rates of substitution can suppress the K_A to K_S ratio below 1.0 even when positive selection has occurred[106,109]. Sequence asymmetry in non-synonymous (but not in synonymous) substitutions between paralogues without $K_A > K_S$ has also been taken as evidence of neofunctionalization[110], but is in fact compatible with other models of duplicate divergence including subfunctionalization[70]. As adaptive evolution can also occur during subfunctionalization (the escape from adaptive conflict model)[60], it is clear that although patterns of sequence evolution provide insights into duplicate gene evolution, they are insufficient to ascertain the mode of divergence.

of ways, not solely to fix or optimize novel functions. The models below give some perspective on the question of selection for novel functions by also addressing other possible fates for duplicated genes.

Neofunctionalization. The classical neofunctionalization model, mutation during non-functionality[10,50] (MDN; FIG. 1a), specifies that a mutation conferring a new function on one member of a duplicate gene pair occurs after the duplication[51]. Here, duplicate fixation can occur either by directional selection (that is, the duplication is fixed because it is beneficial; FIG. 1a) or simply by genetic drift. Note that although it might seem that cases of MDN could be identified based solely on the signature of directional selection in the relevant sequences, there are a number of reasons why this approach can be misleading (BOX 1).

Cases of duplicate gene pairs that show divergent functions, one of which seems to have been gained after the duplication (inferred by phylogenetic comparisons), are plausible examples of this MDN model. A possible instance of MDN is seen in a retrotransposed duplicate copy of the glutamate dehydrogenase gene in humans and apes (*GLUD2*), in which specific changes in allosteric sensitivity and optimal pH seem to be adaptive in its new and restricted expression environment in neurons[52]. *GLUD2* also shows evidence of adaptive evolution after duplication, in contrast to its parent gene, *GLUD1* (REF. 53). These points are strong evidence of the adaptive specialization after duplication that is expected under the MDN model. In this example the adaptation seems to have been initiated by a change in gene expression in the new gene, that is, specialization of expression of *GLUD2* primarily to the brain and testis[54]. Functional changes and directional selection are also evident in the chimeric *jingwei* gene in *Drosophila* that codes for an alcohol dehydrogenase enzyme that has a preference for long-chain alcohols, with a reduced ability to catalyse the oxidation of ethanol — the opposite of the preferences of its parent gene, *Adh*[55]. A third example of functional innovation after gene duplication is the gonadal paralogue of the pig cytochrome P450arom gene. The cytochrome P450 family of proteins generally synthesize oestrogens from androgens. The paralogue is able to synthesize 1 beta-hydroxytestosterone[56], an activity that is absent from paralogues of this gene in pigs and from orthologues in cows and humans (in which 1 beta-hydroxytestosterone cannot be detected[56]).

These three examples highlight an important point: even in cases in which the appearance of the novel function plausibly post-dates the duplication, the duplication is by no means random. Instead, the gene that is eventually appropriated for adaptation after duplication is one with a function that is closely related to the function required. Of course, the mutation process itself remains random, but there is nonrandomness in the selective preservation of useful duplications. Thus, in each case above, novelty is built on a conserved enzymatic mechanism.

Subfunctionalization. Several authors[49,57] have introduced a particular version of model — now known as subfunctionalization — wherein duplicate genes are preserved by purifying selection because the functions of an ancestrally multifunctional gene have become divided up neutrally among the daughter copies. For example, the duplicate gene pair of *SIR3* and *ORC1* was created by the WGD in *S. cerevisiae*. These genes now have divergent functions, in gene silencing and as part of the origin recognition complex, respectively. However, van Hoof[58] showed that an ancestral-type protein from the yeast *Saccharomyces kluyveri*, which lacks the WGD, is able to functionally replace both *S. cerevisiae* genes. This implies that no change in function needs to be invoked to explain the survival of this duplicate pair (see below for a caveat to this analysis).

Subfunctionalization is in fact an older and more general idea[9,37,50] and can occur by two distinct routes. In the route described above, termed the duplication, degeneration, complementation (DDC) model (FIG. 1b), the mutations that cause subfunctionalization are explicitly neutral. In DDC, a gene pair can become preserved in a genome purely as a result of mutations that remove different subsets of the original functions from each gene copy. Each of these mutations is neutral, not deleterious, because the function is still performed by the other copy of the gene. DDC is an attractive explanation for the high frequency of duplicated genes in eukaryotic genomes because it requires only a progenitor gene with more than one function, and ordinary degenerative mutations.

In the other route, termed the escape from adaptive conflict (EAC) model[59–61] (FIG. 1c), the subfunctionalization process involves adaptive (non-neutral) mutations. EAC will occur if the two functions of the ancestral gene cannot simultaneously be optimized by natural selection. After gene duplication, the two daughter genes can escape from the conflict, with each daughter undergoing adaptive mutations that cause it to become specialized towards one of the original subfunctions of the gene. Note that we use 'subfunctionalization' as an umbrella term that covers both the DDC and EAC models.

Retrotransposed
Describes a gene that has undergone duplication through a process that involves an mRNA intermediate. It occurs when a reverse transcriptase enzyme synthesizes DNA from an mRNA template and the DNA is then integrated into the genome. Because retrotransposition usually uses mature mRNAs as a substrate, the resulting duplicate genes often lack introns.

This usage differs from some of the previous literature on the topic in which DDC and subfunctionalization have often been used synonymously, whereas we regard DDC as a particular type of subfunctionalization. Note also that the functions that are partitioned during subfunctionalization can take many forms, for example, gene expression in particular tissues[49], structural domains of genes[62] or quanta of gene expression[63].

Distinguishing between the EAC and DDC models can be difficult, because both show subdividing of ancestral functions. For instance, the *SIR3–ORC1* pair is a clear case of subfunctionalization, but the experimental data are not informative about whether DDC (as reported by the author[58]) or EAC occurred. Distinguishing EAC from MDN can also be difficult, because both exhibit hallmarks of adaptive evolution. Des Marais and Rausher have recently pointed out that one potentially diagnostic feature of EAC is the presence of adaptive evolution in both members of a duplicate gene pair. These authors studied two consecutive duplications, giving a total of three genes, in the dihydroflavonol reductase enzymes in the morning glory family of plants[61]. They found evidence for strong directional selection just before the second duplication, the products of which seem to have lost many of their ancestral enzymatic activities. The third gene, which branched off before this duplication, shows little evidence of selection, but it does show a much higher activity on five substrates compared with two genes tested from taxa without the duplication, which are assumed to represent the ancestral condition. Thus, the authors argue that these duplications allowed specialization of duplicate genes for two sets of functions: the enzymatic activities that they tested and another function that is not yet identified. As the authors note, because we do not know that second function, we cannot show that it also existed in the ancestral gene, meaning that the case for EAC over MDN, although strong, is not yet complete. These authors also suggest that EACs might well be common but unrecognized: later in this Review we describe a pair of duplicates (*GAL1* and *GAL3*) that were previously thought to have diverged by MDN, but that now seem to be a case of EAC.

Interestingly, we also see patterns reminiscent of subfunctionalization in studies of biological networks, although it is not possible to say whether they are of the EAC or DDC type. Several studies have suggested that the extant features of networks, such as those of protein–protein interactions and transcriptional regulation, are compatible with a node (gene) duplication model[64–66]. Notably, however, these models only reproduce the observed degree distributions when duplications are followed by pruning of duplicated interactions[67,68]. This type of subfunctionalization-like pattern is particularly easy to visualize in the context of a WGD (FIG. 2). Indeed, in a study of the particular effects of WGD on network architecture in yeast, we have found evidence for a partitioning of gene expression among duplicate genes produced by WGD. In particular, it seems that one member of a paralogue pair created by WGD is often assigned to stress response pathways[69], whereas the other paralogue does not act in these pathways.

Degree distribution
The degree of a node in a network (in this case, a gene) is the number of interactions it has with other nodes in the network. Thus, in a protein–protein interaction network, the degree of a gene is the number of proteins that the product of the gene interacts with. The degree distribution of a network describes the frequency of nodes in that network with a given degree: many networks of biological interest show a power-law degree distribution.

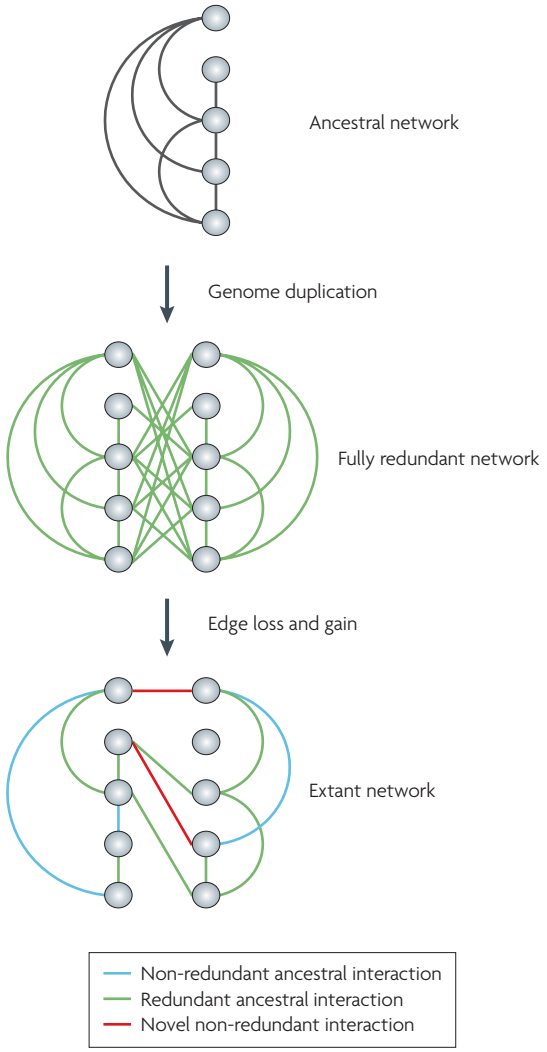


Figure 2 | **Hypothetical example of network evolution following a genome duplication.** Proteins are represented as circles, with interactions between a pair of proteins represented as joining lines. Following genome duplication the number of interactions between proteins is transiently quadrupled, after which a process of interaction loss simplifies the network. Interactions can also be gained during this time, but this process is generally considered to be rarer. Note that we have assumed that the ancestral network is known (which is generally not true for real networks), a fact that allows us to distinguish interactions surviving from the duplication from novel interactions. Figure is modified from REF. 69.

Dosage. Duplicate gene pairs can also be preserved by selection for increased gene dosage[70] (FIG. 1d). Because dosage selection allows duplications to be selectively advantageous at birth, they can be fixed by positive selection and not by genetic drift[70,71]. Examples of potential dosage selection include the *mdr1* gene in the malaria parasite *Plasmodium falciparum*[72], the *CCL3L1* gene in humans[73] and a B esterase gene in mosquitoes[74]. The first two studies associate increased gene copy number with a phenotype of presumably higher fitness (drug or infection resistance)[72,73], whereas the

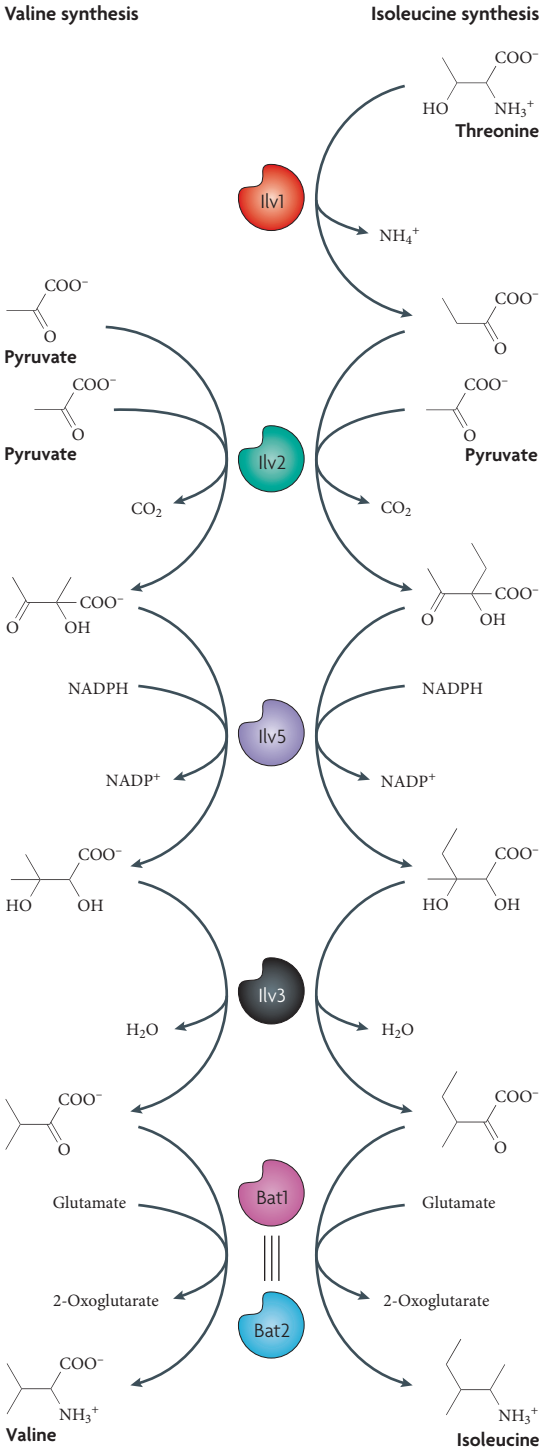


Figure 3 | **The isoleucine and valine biosynthetic pathways in *Saccharomyces cerevisiae*.** Several enzymes catalyse analogous chemical reactions in the two pathways, with similar but distinct substrates. Note that for the step catalysed by Ilv2 (acetolactate synthase) the regulatory subunit Ilv6 is omitted. The triple lines joining proteins Bat1 and Bat2 indicate that these two enzymes are produced by a pair of duplicate genes from the genome duplication in *S. cerevisiae* (>77% amino acid identity[116]). Figure is based on information from REF. 79 and the *Saccharomyces* Genome Database[117].

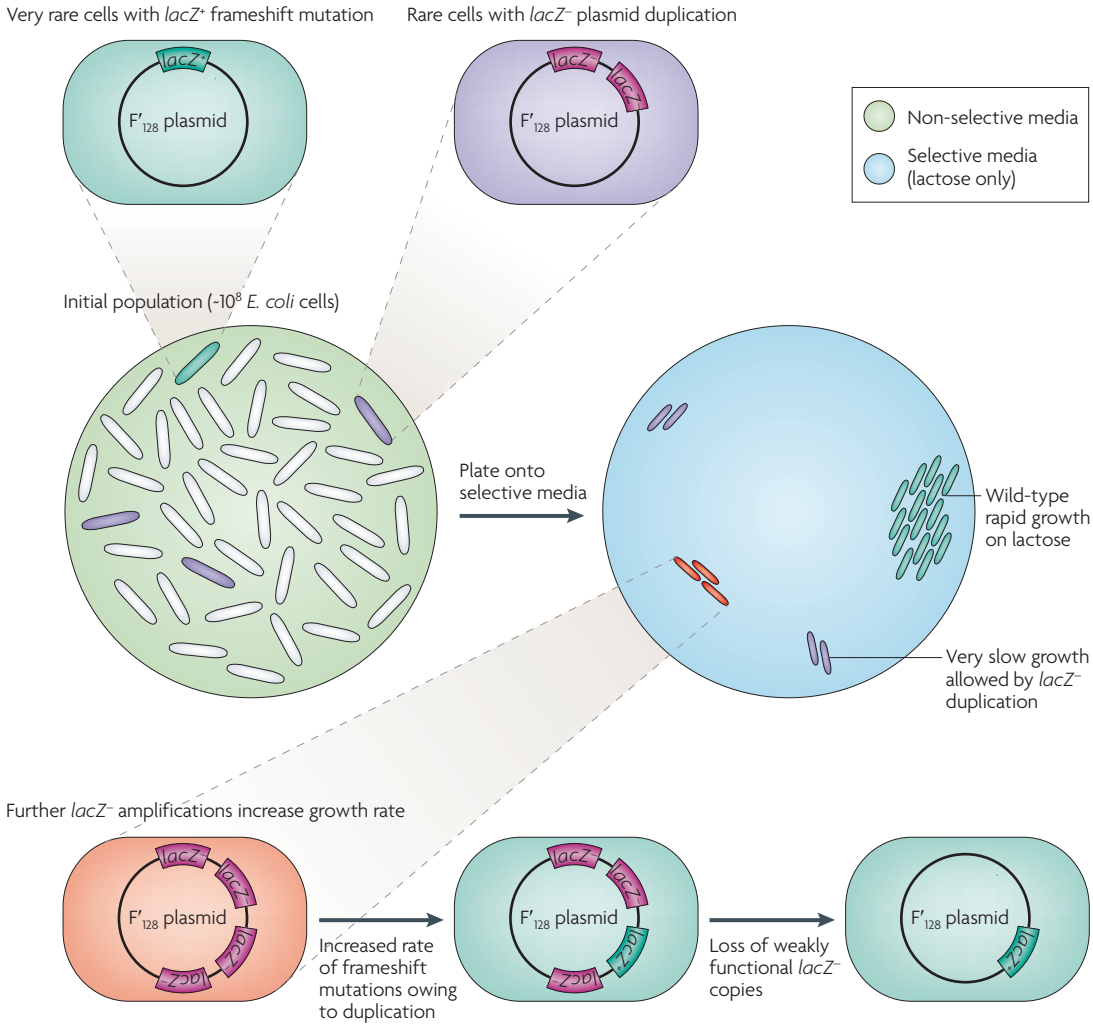
study of the mosquito B esterase gene found that an insecticide challenge results in higher copy numbers of the gene in question[74]. Dosage selection can also be observed directly by using microorganisms in artificial evolution experiments[75–77]. As with subfunctionalization, selection for higher gene dosage might increase the likelihood of a later neofunctionalization event by prolonging the half-life of a duplicate gene pair. However, the above examples also indicate that dosage selection can itself confer novel phenotypes: indeed, dosage selection is a key component of a new model of the origins of novel functions in duplicate genes — the innovation, amplification and divergence (IAD) model.

IAD: building new functions by dosage selection. Genes can be co-opted not only for their primary (evolved) role but also for any of their minor activities. Therefore, the potential range of functions that can be co-opted in a given gene is broad. The notion that enzymes catalyse unique reactions is a useful model of biochemistry, but it has long been known[78] that the real situation is more complicated. For instance, the pathway for isoleucine synthesis is shared with that of valine synthesis in numerous organisms[79] (FIG. 3; other examples are given in REF. 80). Such substrate ambiguity is potentially a vital source of raw material for evolution[78], and the tendency for proteins to have minor functions, or even functions unrelated to their evolved roles, is surprisingly common[81–83]. Examples include the *Escherichia coli* phosphoesterase protein BAP, which can also catalyse the oxidation of phosphite to phosphate[84]. Such multiple functions are not limited to enzymes: at least one known antibody possesses several binding-site conformations, each specific to a different antigen[85]; and below we discuss an example of a steroid receptor with affinity for multiple ligands.

These minor activities can also evolve into new functions. Using a strain of *E. coli* that cannot grow on glucose minimal media owing to the absence of a functional glucokinase gene (*glk*), Miller and Raines[86,87] identified four other genes encoding proteins that can phosphorylate glucose and, when overexpressed, allow growth on glucose minimal media. Thus, these four genes can functionally replace the missing *glk* gene, despite their low sequence identity to it (<25%). In this way, promiscuous protein activities might provide a reservoir of functional novelty. This idea has been combined with an appreciation for the importance of selection for gene dosage into an attractive model of neofunctionalization, which is variously known as the adaptive amplification[88,89], adaptive radiation[90] or IAD[45] model. This model was inspired by the surprising ability of *E. coli* cells to generate 'new' mutations to compensate for the loss of a crucial gene (BOX 2).

The IAD model provides a general mechanism for the evolution of new protein functions[45,90]. Essentially, duplications are selected in certain genes with weak but beneficial activities. The tandem arrays of duplicates that this selection maintains have an increased rate of beneficial, activity-enhancing mutations owing to the

Box 2 | **Origins of the innovation, amplification and divergence model**



Cairns and co-workers created a strain of *Escherichia coli* cells that are able to generate certain kinds of adaptive mutations at very high frequency[111,112]. These cells possessed a 'leaky' knockout of the *lacZ* gene (which encodes β-galactosidase), which was made by eliminating the chromosomal *lac* operon and replacing it with a plasmid containing a frameshifted copy of *lacZ*. The result was a very low, but non-zero, β-galactosidase activity[113]. Populations of such cells cannot grow on media in which lactose is the sole carbon source; this is in part due to the presence of competing scavenger cells[112]. However, under these circumstances, revertants to the wild-type *lacZ* activity were found at rates roughly 100-fold higher than when the selective pressure is absent, a result that did not seem to fit with the standard view of the undirected nature of mutations in evolution. Thus, the result was initially attributed to an increased local mutation rate in *lacZ*[111,112].

However, further enquiry demonstrated a mechanism that did not require the organism to 'know' where and when mutations were required. First, it was noted that many of the revertant colonies had increased copy numbers of the weakly functional *lacZ* gene[88]. When the timing of revertant appearance was analysed, Hendrickson *et al.* found a small class of early revertants that were stable — the *lacZ*$^+$ phenotype was not lost when selection was relaxed[88]. The much larger group of later revertants tended to lose the *lacZ*$^+$ phenotype easily. These observations support a model whereby the early revertants are rare *lacZ* frameshift mutants with full β-galactosidase activity, whereas later revertants restore activity with large arrays of duplicated copies of the weakly functional *lacZ*. Because these large repeats are prone to collapse by recombination in the absence of selection for *lacZ* activity, the phenotype is unstable[88,113].

Hendrickson *et al.* further argue that a duplication of the leaky *lacZ*$^-$ allele persists at low frequency in the initial population (before it is challenged in the lactose-only media)[88]. A single copy of the *lacZ*$^-$ gene is not sufficient to allow growth under these conditions, but this duplication allows slow growth. Selection then operates to favour further duplications in the offspring of these cells (importantly, the duplication rate is enhanced by the presence of the initial tandem duplication[114,115]). Gene amplification continues, increasing the mutational target for the frameshift mutation, because a mutation in any of the copies of *lacZ*$^-$ can yield the wild-type protein. As a result, these amplification-carrying colonies more often generate *lacZ*$^+$ reversions. When this occurs, selection shifts and favours a reduction of the gene array, yielding a single wild-type *lacZ*$^+$ gene[89] (see figure).

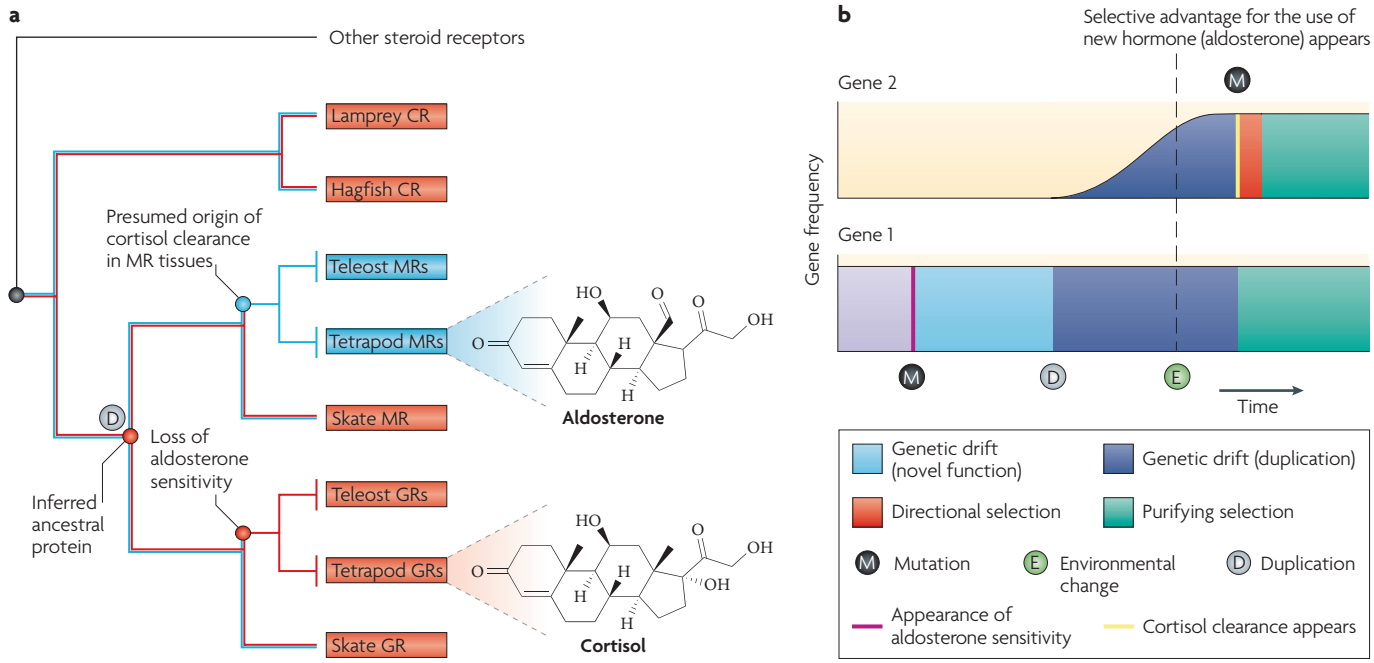


Figure 4 | **Divergence of the aldosterone and cortisol receptors in tetrapods. a** | The aldosterone and the cortisol receptor genes originated by gene duplication, before the split between cartilaginous fish and the lineage leading to tetrapods (point D). This ancestral receptor was recreated in the laboratory[96] and was found to be sensitive to both aldosterone and cortisol, even though there is no evidence for the presence of aldosterone in extant taxa other than tetrapods. It is thus assumed that a non-selected sensitivity was later co-opted: an event that probably required both the loss of aldosterone sensitivity in the cortisol receptor and the reduced use of cortisol in tissues where aldosterone signalling occurs. **b** | A schematic view of one possible version of the timing of events in this example. The secondary aldosterone sensitivity appears early and drifts in the population (light blue region). A subsequent gene duplication also drifts, until a chance mutation specializes one of the copies (yellow line). This new variant is fixed by directional selection, at which point both copies of the gene are under the protection of purifying selection. CR, cortisol receptor; GR, glucocorticoid receptor; MR, mineralcorticoid receptor.

large number of mutational targets. Once such mutations have been fixed, selection to maintain the duplicated array is reduced and the extra copies are lost, leaving only the ancestral gene and the new duplicate with a novel function. The IAD model was originally proposed for microorganisms and, given the initially small selective benefits imparted by the secondary activities of genes, it remains to be seen whether the less efficient action of natural selection in smaller populations will preclude this sort of functional innovation in multicellular organisms.

Beyond simplistic models. The schematics shown in FIG. 1 might suggest that there is a simple taxonomy of fates for duplicate genes. But we prefer to think about the evolution of duplications by considering the timing of the various mutational and selective events. Thus, we see that functional divergence can occur before gene duplication (EAC) or after (MDN), while degenerative substitutions are also occurring independently. Therefore, although neofunctionalization and subfunctionalization are often presented as mutually exclusive alternative fates for duplicate gene pairs, this dichotomy is only valid when considering the mechanism of preservation[12,49,91]. After preservation, duplicate genes continue to evolve, meaning that subfunctionalization can contribute to novelty simply by enabling duplicate genes to survive for long periods, increasing the chances of a neofunctionalizing mutation[92,93].

Of course, considering the timing of events before and after duplication reminds us that these models are conceptual: we will not always have sufficient information to fit real genes from real organisms neatly into one and only one of these categories. It is also worth considering that other, as yet unidentified mechanisms of duplicate gene preservation and adaptation might exist.

One example of co-option of an existing minor function (a 'hobby'), in which the mechanism by which natural selection acted to fix the duplication remains unclear, is found in vertebrates, in the steroid hormone receptor family. The diversity of steroid hormone receptors in teleosts and tetrapods seems to have been built upon an ancestral oestrogen signalling system. When an ancestral steroid receptor was resurrected, it was found to be most responsive to oestrogen compounds and much less responsive to other modern steroids, including progesterone, testosterone, and cortisol[94]. Because progesterone and testosterone are both intermediates in the pathway of oestrogen synthesis, it is reasonable to argue that these intermediates were later co-opted as hormones[95].

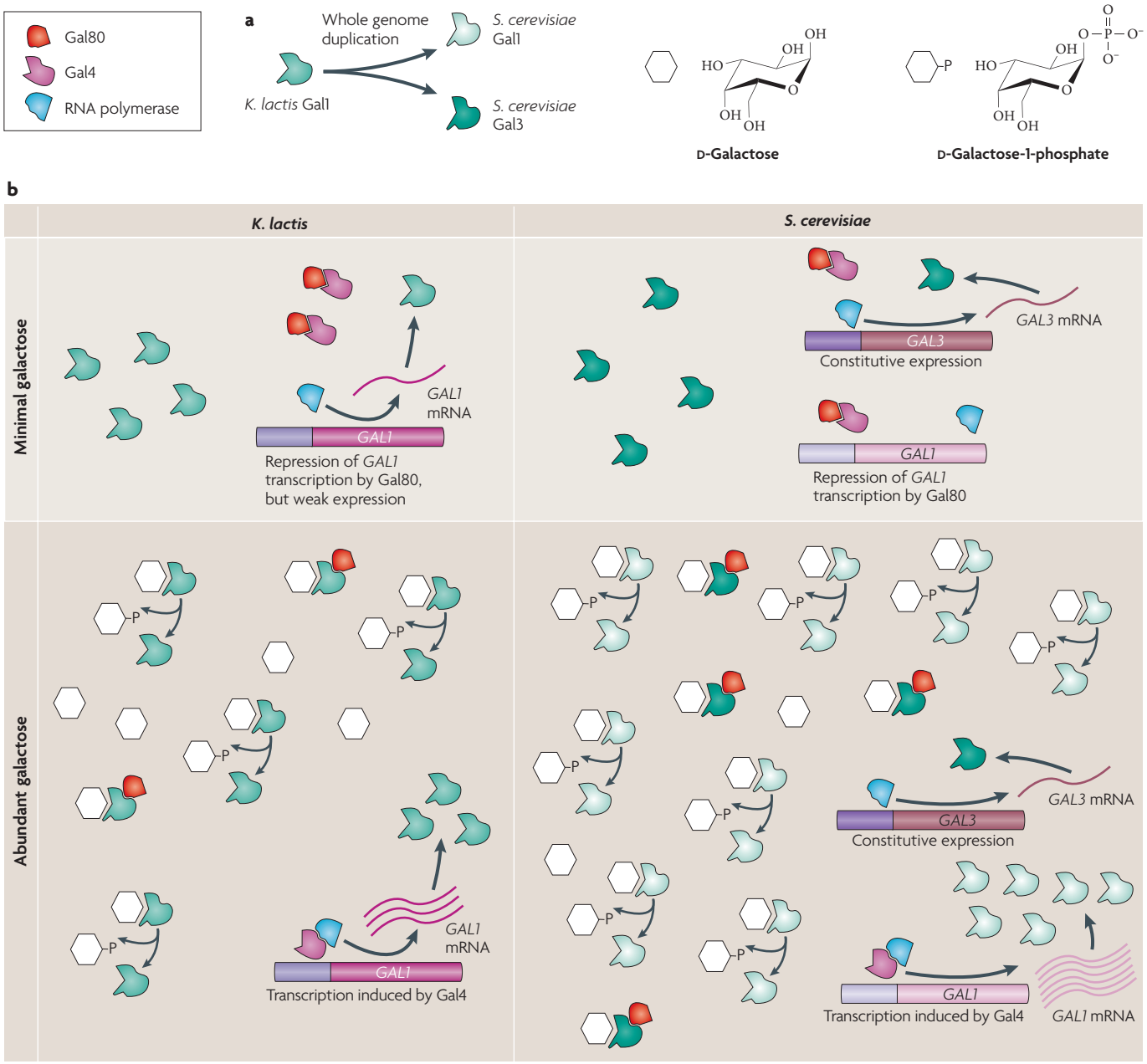


Figure 5 | **Subfunctionalization of a galactose catabolism gene regulatory circuit in *Saccharomyces cerevisiae*.** The ancestral situation is assumed to be represented by the modern yeast *Kluyveromyces lactis*. In this species, a single protein, Gal1 (mid-green), has a dual role as an enzyme and as a regulatory protein. Thus, it can sequester the repressor Gal80 or phosphorylate galactose. The first activity is galactose-dependant, raising the expression of the *GAL1* gene when galactose is present (lower panel). A side-effect of this dual-function role is a lowered dynamic range of transcription, owing to the same promoter being used for the two functions. After the genome duplication of *S. cerevisiae*, the products of the resulting two genes (Gal1 in pale green and Gal3 in dark green) specialized into an enzyme and a regulatory protein, respectively. Thus, Gal3 only binds the repressor whereas Gal1 only phosphorylates galactose. Changes in promoter structure allowed constitutive expression of the *GAL3* gene and very tight coordination between *GAL1* expression and the galactose concentration in the cell[59].

Bridgham *et al.*[96] studied two of these steroid receptors, the mineralcorticoid receptor (MR) and the glucocorticoid receptor (GR), to understand how they diverged in function. The genes encoding these two receptors were formed by duplication at the base of the jawed vertebrate lineage (FIG. 4) and they show distinct specificities, expression patterns and knockout effects in tetrapods[97]; MR receptors bind the hormone aldosterone, whereas GR receptors bind cortisol[98]. Bridgham *et al.*[96] inferred and synthesized the common ancestor of the MR and GR proteins and found that it could be activated by either cortisol or aldosterone. Because

there is no evidence for the presence of aldosterone outside the tetrapods (in which aldosterone production seems to have arisen as a secondary activity in a cortisol-synthesizing enzyme[99]), Bridgham *et al.* argue that the binding of aldosterone by the MR receptor was non-adaptive in origin but was later co-opted by tetrapods after these taxa gained the ability to synthesize aldosterone (an exaptation)[96].

As noted above, for putative cases of MDN it is not always possible to rule out the prior appearance of a novel function in a single gene that subsequently duplicated and diverged, which would make it a case of EAC rather than MDN. Indeed, just this sort of difficulty arose in studies of another example of functional innovation and gene duplication: the *S. cerevisiae* WGD-produced paralogues *GAL1* and *GAL3*, which function as a sugar kinase and as a transcriptional inducer, respectively[100]. One might assume that these two genes diverged in function after duplication (MDN). However, in the related yeast *Kluyveromyces lactis*, the orthologous and unduplicated gene (*GAL1*) serves both functions[101] (FIG. 5), indicating that in *S. cerevisiae GAL1* and *GAL3* are more likely to have diverged through subfunctionalization. Using competition experiments, Hittinger and Carroll[59] showed that the promoter of *GAL1* in *K. lactis* suffers from an adaptive conflict. In the presence of galactose it would be desirable for this organism to induce *GAL1* expression more strongly, and hence to increase the rate of glycolysis. In principle, this could be achieved by increasing the affinity of the promoter for the transcription activator Gal4. However, such promoter mutations are deleterious because they also shut off expression of the gene in the absence of galactose, when the Gal4–Gal80 dimer is a strong repressor, rendering the cell unable to sense galactose if it later becomes available. Gene duplication allowed the specialization of the paralogous promoters in *S. cerevisiae*, making the *GAL3* promoter essentially constitutive while allowing the *GAL1* promoter to be more highly activated by Gal4. The result is both lower *GAL1* expression in the absence of galactose and higher expression in its presence[59].

GAL1 and *GAL3* provide an instructive example of how an adaptive subfunctionalization event (through EAC) can yield a dramatic change in gene function. Thus, Gal1 is an enzyme but not a regulator of gene expression, and Gal3 is a regulator but not an enzyme. This change is adaptive[59], and we can presume that the ability to regulate the genes for galactose catabolism more tightly is useful in ecological niches where this sugar is intermittently abundant. The gene duplication was therefore the ultimate step in a process that converted a galactose-binding protein from an enzyme into a co-activator of transcription. However, even though the argument for the EAC model in Gal1 and Gal3 evolution seems to be strong, much about the innovation is still unknown, including how the enzyme was originally 'captured' for its regulatory role. The *GAL* case is also a noteworthy example of a novel phenotype that primarily owes its origins to a change in gene regulation.

Conclusions

Darwin[1] famously wrote: "If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down." To what degree the modifications of gene function we have described can be seen as slight the reader is left to judge. But Darwin's larger insight was that even 'new' features created by evolution bear the marks of their ancestry, and the key role of co-option in creating new molecular functions exemplifies this principle. In line with Darwin's hypothesis of numerous successive changes, the gene duplications discussed here generally make up only pieces of the story of the adaptation of an organism to its environment.

Achieving comprehensive answers to the twin questions of the origins of novel genes and their refinement by natural selection will require insights to be pooled from genomics, systems biology, population genetics, molecular biology and biochemistry. Such work has already identified new and important questions about duplication and divergence; a lot has been done, but more remains to be done. First, in order to understand which genes are or are not likely to be co-opted to new functions, we need to discover experimentally whether certain gene duplications are actually detrimental at birth and, if so, why. Second, we still do not understand the mechanism by which duplicate genes lose their interactions so soon after duplication; plausible arguments can be made for the action of both genetic drift and natural selection. Finally, we need to know if the reduction of pleiotropic constraints following duplication (which is also evident in this same loss of interactions) is a common event. This reduction could be involved in generating novelty in two ways. It could, as in the *GAL1–GAL3* example, allow the fine-tuning of an existing adaptation. However, it could also provide for future adaptability, although of course selection cannot directly act for such future benefits. Because these last two questions address mechanisms by which duplication might promote the future evolvabilty of an organism, they highlight additional, indirect contributions of duplication to functional innovation.

It is hardly surprising that co-option has a key role in the evolution of new functions through gene duplication. The role of co-option does, however, have implications for the study of duplicate genes. One of the most important implications is that in order to understand the impact of gene duplications (particularly ancient ones) on phenotypes we need to make reference to the biochemistry of the protein as well as to the biology of the organism. And although it is easiest to visualize co-option in the framework of an enzyme that has low levels of activity on secondary substrates, the principle can potentially apply to any protein that interacts with more than one molecule. Our inclination is to categorize the most frequent and kinetically favourable interactions as the right ones and the minor ones as tolerable errors, but in the absence of any grand designer[102] there are no right or wrong interactions — just handholds of different sizes that selection can use to climb a fitness mountain.

1. Darwin, C. *The Origin of Species by Means of Natural Selection* (John Murry, London, 1859).
2. Ruse, M. Is evolution a secular religion? *Science* **299**, 1523–1524 (2003).
3. Müller, G. B. & Wagner, G. P. Novelty in evolution: restructuring the concept. *Annu. Rev. Ecol. Syst.* **22**, 229–256 (1991).
4. Gould, S. J. in *The Panda's Thumb* 19–26 (W. W. Norton, New York, 1980).
5. Gould, S. J. & Vrba, E. S. Exaptation — a missing term in the science of form. *Paleobiology* **8**, 4–15 (1982).
6. Ji, Q., Norell, M. A., Gao, K.-Q., Ji, S.-A. & Ren, D. The distribution of integumentary structures in a feathered dinosaur. *Nature* **410**, 1084–1088 (2001).
7. Xu, X., Zhou, Z.-H. & Prum, R. O. Branched integumental structures in *Sinornithosaurus* and the origin of feathers. *Nature* **410**, 200–204 (2001).
8. Shubin, N., Tabin, C. & Carroll, S. Fossils, genes and the evolution of animal limbs. *Nature* **388**, 639–648 (1997).
9. Taylor, J. S. & Raes, J. Duplication and divergence: the evolution of new genes and old ideas. *Annu. Rev. Genet.* **38**, 615–643 (2004).
10. Ohno, S. *Evolution by Gene Duplication* (Springer, New York, 1970).
11. Lynch, M. The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl Acad. Sci. USA* **104** (Suppl. 1), 8597–8604 (2007).
12. Lynch, M. & Conery, J. S. The evolutionary fate and consequences of duplicate genes. *Science* **290**, 1151–1155 (2000).
A landmark paper that was the first to estimate genome-wide rates of formation and death of duplicated genes in eukaryotes, and the first to demonstrate the relaxation of selective constraints on duplicated genes.
13. Davis, J. C. & Petrov, D. A. Preferential duplication of conserved proteins in eukaryotic genomes. *PLoS Biol.* **2**, E55 (2004).
14. Li, L., Huang, Y., Xia, X. & Sun, Z. Preferential duplication in the sparse part of yeast protein interaction network. *Mol. Biol. Evol.* **23**, 2467–2473 (2006).
15. He, X. & Zhang, J. Higher duplicability of less important genes in yeast genomes. *Mol. Biol. Evol.* **23**, 144–151 (2006).
16. Deluna, A. *et al.* Exposing the fitness contribution of duplicated genes. *Nature Genet.* **40**, 676–681 (2008).
17. Drummond, D. A., Raval, A. & Wilke, C. O. A single determinant dominates the rate of yeast protein evolution. *Mol. Biol. Evol.* **23**, 327–337 (2006).
The authors demonstrate that the rate of translation of mRNA is the major determinant of nonsynonymous and synonymous evolutionary rates in yeasts.
18. Paterson, A. H. *et al.* Many gene and domain families have convergent fates following independent whole-genome duplication events in *Arabidopsis*, *Oryza*, *Saccharomyces* and *Tetraodon*. *Trends Genet.* **22**, 597–602 (2006).
19. Papp, B., Pal, C. & Hurst, L. D. Dosage sensitivity and the evolution of gene families in yeast. *Nature* **424**, 194–197 (2003).
20. Wapinski, I., Pfeffer, A., Friedman, N. & Regev, A. Natural history and evolutionary principles of gene duplication in fungi. *Nature* **449**, 54–61 (2007).
21. Aury, J. M. *et al.* Global trends of whole-genome duplications revealed by the ciliate *Paramecium tetraurelia*. *Nature* **444**, 171–178 (2006).
22. Seoighe, C. & Wolfe, K. H. Yeast genome evolution in the post-genome era. *Curr. Opin. Microbiol.* **2**, 548–554 (1999).
23. Blanc, G. & Wolfe, K. H. Functional divergence of duplicated genes formed by polyploidy during *Arabidopsis* evolution. *Plant Cell* **16**, 1679–1691 (2004).
24. Harbison, C. T. *et al.* Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
25. Maere, S. *et al.* Modeling gene and genome duplications in eukaryotes. *Proc. Natl Acad. Sci. USA* **102**, 5454–5459 (2005).
The first application of mathematical modelling techniques to study variation in the propensities of different functional classes of genes to survive in duplicate after WGD or SSD events.
26. Guan, Y., Dunham, M. J. & Troyanskaya, O. G. Functional analysis of gene duplications in *Saccharomyces cerevisiae*. *Genetics* **175**, 933–943 (2007).
27. Hakes, L., Pinney, J. W., Lovell, S. C., Oliver, S. G. & Robertson, D. L. All duplicates are not equal: the difference between small-scale and genome duplication. *Genome Biol.* **8**, R209 (2007).
28. Scannell, D. R. *et al.* Independent sorting-out of thousands of duplicated gene pairs in two yeast species descended from a whole-genome duplication. *Proc. Natl Acad. Sci. USA* **104**, 8397–8402 (2007).
29. Conant, G. C. & Wolfe, K. H. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol. Syst. Biol.* **3**, 129 (2007).
30. Freeling, M. & Thomas, B. C. Gene-balanced duplications, like tetraploidy, provide predictable drive to increase morphological complexity. *Genome Res.* **16**, 805–814 (2006).
31. Dean, E. J., Davis, J. C., Davis, R. W. & Petrov, D. A. Pervasive and persistent redundancy among duplicated genes in yeast. *PLoS Genet.* **4**, e1000113 (2008).
32. Musso, G., Zhang, Z. & Emili, A. Retention of protein complex membership by ancient duplicated gene products in budding yeast. *Trends Genet.* **23**, 266–269 (2007).
33. Byrne, K. P. & Wolfe, K. H. Consistent patterns of rate asymmetry and gene loss indicate widespread neofunctionalization of yeast genes after whole-genome duplication. *Genetics* **175**, 1341–1350 (2007).
34. Piškur, J., Rozpedowska, E., Polakova, S., Merico, A. & Compagno, C. How did *Saccharomyces* evolve to become a good brewer? *Trends Genet.* **22**, 183–186 (2006).
35. Thomson, J. M. *et al.* Resurrecting ancestral alcohol dehydrogenases from yeast. *Nature Genet.* **37**, 630–635 (2005).
36. Zhang, J., Zhang, Y. P. & Rosenberg, H. F. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. *Nature Genet.* **30**, 411–415 (2002).
37. Piatigorsky, J. *et al.* Gene sharing by delta-crystallin and argininosuccinate lyase. *Proc. Natl Acad. Sci. USA* **85**, 3479–3483 (1988).
38. Wistow, G. & Piatigorsky, J. Recruitment of enzymes as lens structural proteins. *Science* **236**, 1554–1556 (1987).
39. Chen, L., DeVries, A. L. & Cheng, C. H. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. *Proc. Natl Acad. Sci. USA* **94**, 3811–3816 (1997).
40. Smith, A. A. *et al.* Gene duplication and separation of functions in alphaB-crystallin from zebrafish (*Danio rerio*). *FEBS J.* **273**, 481–490 (2006).
41. Williford, A., Stay, B. & Bhattacharya, D. Evolution of a novel function: nutritive milk in the viviparous cockroach, *Diploptera punctata*. *Evol. Dev.* **6**, 67–77 (2004).
42. Korchi, A., Brossut, R., Bouhin, H. & Delachambre, J. cDNA cloning of an adult male putative lipocalin specific to tergal gland aphrodisiac secretion in an insect (*Leucophaea maderae*). *FEBS Lett.* **449**, 125–128 (1999).
43. Lynch, M. & Conery, J. S. The origins of genome complexity. *Science* **302**, 1401–1404 (2003).
44. Wagner, A. Energy constraints on the evolution of gene expression. *Mol. Biol. Evol.* **22**, 1365–1374 (2005).
45. Bergthorsson, U., Andersson, D. I. & Roth, J. R. Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl Acad. Sci. USA* **104**, 17004–17009 (2007).
The authors propose the IAD model of co-option of minor activities of genes.
46. Gerstein, A. C., Chun, H.-J., Grant, A. & Otto, S. P. Genomic convergence toward diploidy in *Saccharomyces cerevisiae*. *PLoS Genet.* **2**, e145 (2006).
47. Li, W.-H. Rate of gene silencing at duplicate loci: a theoretical study and interpretation of data from tetraploid fish. *Genetics* **95**, 237–258 (1980).
48. Nei, M. & Roychoudhury, A. K. Probability of fixation of nonfunctional genes at duplicate loci. *Am. Nat.* **107**, 362–372 (1973).
49. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).
The paper that introduced the terms subfunctionalization and neofunctionalization, and that proposed that a specific form of subfunctionalization (DDC) is a frequent cause of the initial preservation of duplicated genes in eukaryotes.
50. Hughes, A. L. The evolution of functionally novel proteins after gene duplication. *Proc. Roy. Soc. Lond. B* **256**, 119–124 (1994).
51. Kimura, M. & Ohta, T. On some principles governing molecular evolution. *Proc. Natl Acad. Sci. USA* **71**, 2848–2852 (1974).
52. Plaitakis, A., Spanaki, C., Mastorodemos, V. & Zaganas, I. Study of structure–function relationships in human glutamate dehydrogenases reveals novel molecular mechanisms for the regulation of the nerve tissue-specific (*GLUD2*) isoenzyme. *Neurochem. Int.* **43**, 401–410 (2003).
53. Burki, F. & Kaessmann, H. Birth and adaptive evolution of a hominoid gene that supports high neurotransmitter flux. *Nature Genet.* **36**, 1061–1063 (2004).
54. Shashidharan, P. *et al.* Novel human glutamate dehydrogenase expressed in neural and testicular tissues and encoded by an X-linked intronless gene. *J. Biol. Chem.* **269**, 16971–16976 (1994).
55. Zhang, J., Dean, A. M., Brunet, F. & Long, M. Evolving protein functional diversity in new genes of *Drosophila*. *Proc. Natl Acad. Sci. USA* **101**, 16246–16250 (2004).
56. Corbin, C. J. *et al.* Paralogues of porcine aromatase cytochrome P450: a novel hydroxylase activity is associated with the survival of a duplicated gene. *Endocrinology* **145**, 2157–2164 (2004).
57. Stoltzfus, A. On the possibility of constructive neutral evolution. *J. Mol. Evol.* **49**, 169–181 (1999).
58. van Hoof, A. Conserved functions of yeast genes support the duplication, degeneration and complementation model for gene duplication. *Genetics* **171**, 1455–1461 (2005).
59. Hittinger, C. T. & Carroll, S. B. Gene duplication and the adaptive evolution of a classic genetic switch. *Nature* **449**, 677–681 (2007).
60. Hughes, A. L. Gene duplication and the origin of novel proteins. *Proc. Natl Acad. Sci. USA* **102**, 8791–8792 (2005).
61. Des Marais, D. L. & Rausher, M. D. Escape from adaptive conflict after duplication in an anthocyanin pathway gene. *Nature* **454**, 762–765 (2008).
62. Cusack, B. P. & Wolfe, K. H. When gene marriages don't work out: divorce by subfunctionalization. *Trends Genet.* **23**, 270–272 (2007).
63. Blanchard, J. L. & Lynch, M. Organellar genes: why do they end up in the nucleus? *Trends Genet.* **16**, 315–320 (2000).
64. Teichmann, S. & Babu, M. M. Gene regulatory network growth by duplication. *Nature Genet.* **36**, 492–496 (2004).
65. Berg, J., Lässig, M. & Wagner, A. Structure and evolution of protein interaction networks: a statistical model for link dynamics and gene duplications. *BMC Evol. Biol.* **4**, 51 (2004).
66. Chung, F., Lu, L., Dewey, T. G. & Galas, D. J. Duplication models for biological networks. *J. Comput. Biol.* **10**, 677–687 (2003).
67. Wagner, A. How the global structure of protein interaction networks evolves. *Proc. Roy. Soc. Lond. B* **270**, 457–466 (2003).
68. Pastor-Satorras, R., Smith, E. & Sole, R. V. Evolving protein interaction networks through gene duplication. *J. Theor. Biol.* **222**, 199–210 (2003).
69. Conant, G. C. & Wolfe, K. H. Functional partitioning of yeast co-expression networks after genome duplication. *PLoS Biol.* **4**, e109 (2006).
70. Kondrashov, F. A. & Kondrashov, A. S. Role of selection in fixation of gene duplications. *J. Theor. Biol.* **239**, 141–151 (2006).
71. Kondrashov, F. A., Rogozin, I. B., Wolf, Y. I. & Koonin, E. V. Selection in the evolution of gene duplications. *Genome Biol.* **3**, 0008.1–0008.9 (2002).
72. Price, R. N. *et al.* Mefloquine resistance in *Plasmodium falciparum* and increased *pfmdr1* gene copy number. *Lancet* **364**, 438–447 (2004).
73. Gonzalez, E. *et al.* The influence of *CCL3L1* gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* **307**, 1434–1440 (2005).
74. Guillemaud, T. *et al.* Quantitative variation and selection of esterase gene amplification in *Culex pipiens*. *Heredity* **83** (Pt 1), 87–99 (1999).
75. Riehle, M. M., Bennett, A. F. & Long, A. D. Genetic architecture of thermal adaptation in *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **98**, 525–530 (2001).
76. Brown, C. J., Todd, K. M. & Rosenzweig, R. F. Multiple duplications of yeast hexose-transport genes in response to selection in a glucose-limited environment. *Mol. Biol. Evol.* **15**, 931–942 (1998).
77. Sonti, R. V. & Roth, J. R. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. *Genetics* **123**, 19–28 (1989).

78. Jensen, R. A. Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* **30**, 409–425 (1976).
79. Voet, D. & Voet, J. G. *Biochemistry* (John Wiley & Sons, Inc., Hoboken, New Jersey, 2004).
80. Gancedo, C. & Flores, C. L. Moonlighting proteins in yeasts. *Microbiol. Mol. Biol. Rev.* **72**, 197–210 (2008).
81. James, L. C. & Tawfik, D. S. Conformational diversity and protein evolution — a 60-year-old hypothesis revisited. *Trends Biochem. Sci.* **28**, 361–368 (2003).
82. Copley, S. D. Enzymes with extra talents: moonlighting functions and catalytic promiscuity. *Curr. Opin. Chem. Biol.* **7**, 265–272 (2003).
83. O'Brien, P. J. & Herschlag, D. Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **6**, R91–R105 (1999).
84. Yang, K. & Metcalf, W. W. A new activity for an old enzyme: *Escherichia coli* bacterial alkaline phosphatase is a phosphite-dependent hydrogenase. *Proc. Natl Acad. Sci. USA* **101**, 7919–7924 (2004).
85. James, L. C., Roversi, P. & Tawfik, D. S. Antibody multispecificity mediated by conformational diversity. *Science* **299**, 1362–1367 (2003).
The authors describe an antibody that not only binds distinct antigens (a small aromatic compound and a protein) but seems to do so as the result of possessing at least two distinct free-state structural conformations.
86. Miller, B. G. & Raines, R. T. Identifying latent enzyme activities: substrate ambiguity within modern bacterial sugar kinases. *Biochemistry* **43**, 6387–6392 (2004).
87. Miller, B. G. & Raines, R. T. Reconstitution of a defunct glycolytic pathway via recruitment of ambiguous sugar kinases. *Biochemistry* **44**, 10776–10783 (2005).
88. Hendrickson, H., Slechta, E. S., Bergthorsson, U., Andersson, D. I. & Roth, J. R. Amplification-mutagenesis: evidence that 'directed' adaptive mutation and general hypermutability result from growth with a selected gene amplification. *Proc. Natl Acad. Sci. USA* **99**, 2164–2169 (2002).
89. Roth, J. R. & Andersson, D. I. Adaptive mutation: how growth under selection stimulates Lac(+) reversion by increasing target copy number. *J. Bacteriol.* **186**, 4855–4860 (2004).
90. Francino, M. P. An adaptive radiation model for the origin of new gene functions. *Nature Genet.* **37**, 573–577 (2005).
91. Lynch, M. & Force, A. The probability of duplicate gene preservation by subfunctionalization. *Genetics* **154**, 459–473 (2000).
92. He, X. & Zhang, J. Rapid subfunctionalization accompanied by prolonged and substantial neofunctionalization in duplicate gene evolution. *Genetics* **169**, 1157–1164 (2005).
93. Rastogi, S. & Liberles, D. A. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* **5**, 28 (2005).
94. Thornton, J. W., Need, E. & Crews, D. Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling. *Science* **301**, 1714–1717 (2003).
95. Thornton, J. W. Evolution of vertebrate steroid receptors from an ancestral estrogen receptor by ligand exploitation and serial genome expansions. *Proc. Natl Acad. Sci. USA* **98**, 5671–5676 (2001).
96. Bridgham, J. T., Carroll, S. M. & Thornton, J. W. Evolution of hormone-receptor complexity by molecular exploitation. *Science* **312**, 97–101 (2006).
97. Farman, N. & Rafestin-Oblin, M. E. Multiple aspects of mineralocorticoid selectivity. *Am. J. Physiol. Renal Physiol.* **280**, F181–F192 (2001).
98. Gronemeyer, H., Gustafsson, J. A. & Laudet, V. Principles for modulation of the nuclear receptor superfamily. *Nature Rev. Drug Discov.* **3**, 950–964 (2004).
99. Bülow, H. E. & Bernhardt, R. Analyses of the *CYP11B* gene family in the guinea pig suggest the existence of a primordial *CYP11B* gene with aldosterone synthase activity. *Eur. J. Biochem.* **269**, 3838–3846 (2002).
100. Bhat, P. J. & Murthy, T. V. Transcriptional control of the *GAL/MEL* regulon of yeast *Saccharomyces cerevisiae*: mechanism of galactose-mediated signal transduction. *Mol. Microbiol.* **40**, 1059–1066 (2001).
101. Zenke, F. T. *et al.* Activation of Gal4p by galactose-dependent interaction of galactokinase and Gal80p. *Science* **272**, 1662–1665 (1996).
102. Warda, M. & Han, J. Retracted: Mitochondria, the missing link between body and soul: proteomic prospective evidence. *Proteomics* **8**, I–XXIII (2008).
103. Shiu, S. H., Byrnes, J. K., Pan, R., Zhang, P. & Li, W. H. Role of positive selection in the retention of duplicate genes in mammalian genomes. *Proc. Natl Acad. Sci. USA* **103**, 2232–2236 (2006).
104. Nielsen, R. *et al.* A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170 (2005).
105. Marques, A. C., Dupanloup, I., Vinckenbosch, N., Reymond, A. & Kaessmann, H. Emergence of young human genes after a burst of retroposition in primates. *PLoS Biol.* **3**, e357 (2005).
106. Kreitman, M. & Akashi, H. Molecular evidence for natural selection. *Annu. Rev. Ecol. Syst.* **26**, 403–422 (1995).
107. Yang, Z. Maximum likelihood estimation on large phylogenies and analysis of adaptive evolution in human influenza virus A. *J. Mol. Evol.* **51**, 423–432 (2000).
108. Zhang, J. Z., Rosenberg, H. F. & Nei, M. Positive Darwinian selection after gene duplication in primate ribonuclease genes. *Proc. Natl Acad. Sci. USA* **95**, 3708–3713 (1998).
109. Sharp, P. M. In search of molecular darwinism. *Nature* **385**, 111–112 (1997).
110. Chain, F. J. & Evans, B. J. Multiple mechanisms promote the retained expression of gene duplicates in the tetraploid frog *Xenopus laevis*. *PLoS Genet.* **2**, e56 (2006).
111. Cairns, J., Overbaugh, J. & Miller, S. The origin of mutants. *Nature* **335**, 142–145 (1988).
112. Cairns, J. & Foster, P. L. Adaptive reversion of a frameshift mutation in *Escherichia coli*. *Genetics* **128**, 695–701 (1991).
113. Andersson, D. I., Slechta, E. S. & Roth, J. R. Evidence that gene amplification underlies adaptive mutability of the bacterial *lac* operon. *Science* **282**, 1133–1135 (1998).
114. Kugelberg, E., Kofoid, E., Reams, A. B., Andersson, D. I. & Roth, J. R. Multiple pathways of selected gene amplification during adaptive mutation. *Proc. Natl Acad. Sci. USA* **103**, 17319–17324 (2006).
115. Romero, D. & Palacios, R. Gene amplification and genomic plasticity in prokaryotes. *Annu. Rev. Genet.* **31**, 91–111 (1997).
116. Byrne, K. P. & Wolfe, K. H. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res.* **15**, 1456–1461 (2005).
117. Cherry, J. M. *et al.* SGD: *Saccharomyces* Genome Database. *Nucleic Acids Res.* **26**, 73–80 (1998).

Acknowledgements
We thank K. Byrne, B. Cusack, J. Gordon, N. Khaldi, and J. Mower for discussions regarding the fates of duplicated genes. We would also like to thank three anonymous reviewers for critical comments. This work was supported by Science Foundation Ireland.

DATABASES
Entrez Gene: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=gene
GAL1 | *GAL3* | *GLUD1* | *GLUD2*

FURTHER INFORMATION
Gavin C. Conant's homepage: http://web.missouri.edu/~conantg
Wolfe Laboratory homepage: http://wolfe.gen.tcd.ie
PubCrawler: www.pubcrawler.ie

ALL LINKS ARE ACTIVE IN THE ONLINE PDF